TITLE: Template Matching in the l1 Norm

# Template matching in the $\ell_1$ norm

**Iain J. Anderson and Colin Ross**

*School of Computing and Mathematics, University of Huddersfield, UK.*
i.j.anderson@hud.ac.uk, c.ross@hud.ac.uk

**Abstract**

We present a method for matching a surface in three dimensions to a set of data sampled from the surface by means of minimising the distances from the data points to the closest point on the surface. This method of association is affine transformation invariant and as such is very useful in situations where the coordinate axes are essentially arbitrary. Traditionally, this problem has been solved by minimising the $\ell_2$ norm of the distances from the data points to the corresponding points in the surface, while the use of other $\ell_p$ norms is less well known. We present a method for template matching in the $\ell_1$ norm based upon a method of directional constraints developed by Watson for the related problem of orthogonal distance regression. An algorithm for this method is given and numerical results show its effectiveness.

## 1 Introduction

Template matching is used in a variety of applications such as the quality assurance of manufactured artifacts [1] and dental metrology [2]. Given a fixed template, i.e., curve or surface, and a set of data in a different frame of reference, template matching involves finding the frame transformation which maps the data onto the template.

A typical strategy for finding the optimal transformation parameters in the template matching problem is to minimize, in some norm, the orthogonal distances between the transformed data and the template. In this case, the template matching problem can be viewed as a form of orthogonal distance regression (ODR) [3], which is a technique commonly used for fitting curves and surfaces to measured data. Therefore, most algorithms for solving the template matching problem are extensions of algorithms for ODR. Template matching in the $\ell_2$ norm is addressed by Turner [3] and in the $\ell_\infty$ norm by Butler et al. [1] as well as by Zwick [7] for the two dimensional case.

In this paper, we are specifically concerned with the following problem.

> Given a fixed differentiable parametric surface $\mathbf{f}(u, v)$ and a set of $m$ data $\{\mathbf{x}_i\}_{i=1}^m \in \Re^3$, find points $\{\mathbf{f}(u_i, v_i)\}_{i=1}^m$, a rotation matrix $R_\Theta$, and a translation vector $\mathbf{t}_0$ such that the $\ell_1$ norm of the residual distances $\{\|R_\Theta(\mathbf{x}_i - \mathbf{t}_0) - \mathbf{f}(u_i, v_i)\|_2\}_{i=1}^m$ is minimal.

This is the template matching problem in the $\ell_1$ norm, and although not as widely used as the $\ell_2$ and $\ell_\infty$ counterparts, it does nonetheless have an important role to play. The importance of the $\ell_1$ norm is that, generally speaking, any outlying data are effectively ignored with the result that an approximation is obtained which is largely independent

of any unreliable data. This has particular importance when our data arises as a result of some measurement process, perhaps involving many complicated and finely-tuned instruments. For such a measurement scenario, any change in the assumed measurement conditions can result in a datum which has gross error relative to other data. Thus, if we choose a measure which is susceptible to outlying data, we are in danger of obtaining an unrepresentative approximation. This situation is avoided by use of the $\ell_1$ norm and we therefore advocate its use both here and in any situation involving measurement data where a representative approximation is required.

A feature of optimal $\ell_1$ solutions is the likelihood of a small number of the data having a residual of zero, and it is therefore unclear whether the elements of the Jacobian matrix of partial derivatives are well-defined for these points. As a result, use of the usual Gauss-Newton method would appear to be handicapped due to its dependence upon the Jacobian matrix to calculate an updated transformation estimate. This difficulty also arises in the conventional ODR fitting problem and has recently been considered by Watson [6]. His solution is to adopt a method of fitting subject to directional constraints. By setting these directional constraints to be orthogonal to the approximant, Watson shows not only that the Jacobian is defined but also how to compute its elements without incurring a build-up of rounding error.

In this paper, we extend Watson's constrained direction fitting routine to the template matching problem. We show that Watson's results are equally valid for $\ell_1$ template matching. Finally, we exploit these results to give a reliable algorithm for the $\ell_1$ template matching problem.

The structure of this paper is as follows. Section 2 provides the results necessary to justify the new technique. Section 3 describes the algorithm adopted to implement the theory. Section 4 gives some numerical results for both a simple case and a more challenging case. Finally, Section 5 concludes this paper and presents possibilities for future work.

## 2 Theory

We are concerned with the minimisation of the quantity

$$E = \|(d_1, \ldots, d_m)\|, \tag{2.1}$$

where

$$d_i = \min_{u_i, v_i} \|\widehat{\mathbf{x}}_i - \mathbf{f}(u_i, v_i)\|_2, \quad i = 1, 2, \ldots, m, \tag{2.2}$$

and

$$\widehat{\mathbf{x}} = R_\Theta(\mathbf{x} - \mathbf{t}_0), \tag{2.3}$$

with respect to the rotation parameters

$$\Theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix},$$

the translation parameters

$$\mathbf{t}_0 = \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

and the location parameters

$$U = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

This is a constrained problem and can be solved using a separation-of-variables approach as described by Turner [3] among others. In this approach, the problem of obtaining the transformation parameters

$$\mathbf{t} = \begin{pmatrix} \Theta \\ \mathbf{t}_0 \end{pmatrix},$$

is separated from the subproblem of obtaining the location parameters $U$. At each iteration, the subproblem is solved to obtain an optimal $U$ for the current transformation parameters $\mathbf{t}$ which is then used to obtain an update of the transformation parameters themselves.

### 2.1 Considerations specific to the $\ell_1$ problem

Up to this point, we have not specified which norm we are using to measure the disparity between the transformed data and the template. Since we will be particularly interested in the $\ell_1$ case, this section discusses problems inherent in the solution of such a problem.

The major problem with solving non-linear $\ell_1$ problems is that in order to use a technique such as the Gauss-Newton method, derivative information is required. Unfortunately, derivatives of the distances $\mathbf{d}$ are not defined when a distance has a value of zero. Such is the nature of $\ell_1$ approximation that zeros are to be expected at an optimal solution [5]. Thus, it is unclear whether the Jacobian matrix is defined at these data points. Recent work by Watson [6] has considered how the related problem of orthogonal distance regression might be solved by considering distances to be measured along fixed *direction vectors* $\mathbf{w}_i$. Orthogonal distance regression involves the fitting of a curve or surface to a set of data where the residuals are taken to be the shortest distance from the data to the approximant [3]. Template matching can be seen as a variant of this since the residuals are measured in the same way, but we are only altering the position and orientation of the approximant, rather than the actual shape itself. Thus, techniques for orthogonal distance regression can be used successfully in template matching.

By means of these directional constraints, it is possible to show that if we choose the directions $\mathbf{w}_i$ to be the orthogonal directions,

$$\mathbf{w}_i = \frac{\mathbf{f}(u_i, v_i) - \widehat{\mathbf{x}}_i}{\|\mathbf{f}(u_i, v_i) - \widehat{\mathbf{x}}_i\|_2}$$

then the derivatives are well defined in the limit as $\|\mathbf{f}(u_i, v_i) - \widehat{\mathbf{x}}_i\|_2 \to 0$.

This result may be summarised in the following Theorem (taken from Watson [6]).

**Theorem 2.1** *For parametric fitting, let the (usual) Gauss-Newton method produce a sequence* $\{\mathbf{t}\}$ *such there is a unique unit normal vector to the template at* $\mathbf{f}(u_i, v_i)$*, and*

$\widehat{\mathbf{x}}_i$ *remains on one side of the template. Then $\nabla_{\mathbf{t}} d_i$ is well defined on this sequence.*

If $\mathbf{f}(u_i, v_i) \rightarrow \widehat{\mathbf{x}}_i$, then this formulation will lead to similar problems to which we are attempting to resolve as a result of the quotient becoming undefined. As a result, Watson [6] suggests leaving $\mathbf{w}_i$ unchanged once $d_i$ becomes small. By this method, numerical problems arising as a result of a distance tending to zero may be avoided. However, the algorithm will still tend to the correct solution provided that the small residual corresponds to an interpolation point of the $\ell_1$ solution. If this is not the case, then the solution will not be optimal, but will still be close to the optimal solution.

### 2.2 Possible problems

The most immediate problem that arises is how to ensure that there exists a point on the template which is situated along the direction vector given from each datum. Clearly in certain situations, there will not exist such a point — corresponding to the case where the direction vector lies within the tangent plane of the template in the region of the datum. In such a situation there would seem to be two possible recourses available.

(1) Ignore these data.
(2) Choose the point on the template that is closest to the line though the datum defined by the direction vector.

It has been found through empirical results that provided the problem only occurs on certain iterations rather than as a result of poor choice of the direction vectors associated with the template, ignoring the problem data is the better option. Use of the second option has been found to prevent convergence of the algorithm.

## 3 Algorithm

The algorithm to implement this technique consists of two sub-algorithms, each related to a specific section of the main algorithm. These sub-algorithms are

(1) the constrained closest point problem,
(2) the calculation of a new transformation estimate.

### 3.1 Constrained closest point problem

For each data point $\mathbf{x}_i$, this problem is that of finding $u_i$ and $v_i$ such that the constraint

$$\widehat{\mathbf{x}} - \mathbf{f}(u, v) = d\mathbf{w}, \tag{3.1}$$

is satisfied (subscripts dropped for clarity). Expanding this equation, we obtain

$$\begin{pmatrix} \widehat{x} \\ \widehat{y} \\ \widehat{z} \end{pmatrix} - \begin{pmatrix} f \\ g \\ h \end{pmatrix} - d \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = 0.$$

If we pre-multiply this equation by $\mathbf{a}^{\mathrm{T}}$, we obtain

$$\mathbf{a}^{\mathrm{T}}\widehat{\mathbf{x}} - \mathbf{a}^{\mathrm{T}}\mathbf{f}(u, v) - d\mathbf{a}^{\mathrm{T}}\mathbf{w} = 0. \tag{3.2}$$

Thus, by choosing $\mathbf{a}$ to be orthogonal to $\mathbf{w}$, we are able to eliminate $d$ from equation (3.2). Similarly, if we multiply equation (3.1) by $\mathbf{b}$ we obtain the equation

$$\mathbf{b}^{\mathrm{T}}\widehat{\mathbf{x}} - \mathbf{b}^{\mathrm{T}}\mathbf{f}(u,v) - d\mathbf{b}^{\mathrm{T}}\mathbf{w} = 0.$$

We may thereby reduce the system (3.1) to that of two (nonlinear) equations in two unknowns ($u$ and $v$). This system can then be solved by adopting a Newton-type method. Our problem has been reduced to that of solving

$$F(u,v) = [\mathbf{a} : \mathbf{b}]^{\mathrm{T}}(\widehat{\mathbf{x}} - \mathbf{f}(u,v)) = 0,$$

which has derivative

$$\nabla_{u,v}F = -[\mathbf{a} : \mathbf{b}]^{\mathrm{T}}(\nabla_u \mathbf{f} : \nabla_v \mathbf{f}),$$

by means of Newton's method which involves adopting an iterative approach and solving

$$\nabla_{u,v}F \begin{pmatrix} \delta u \\ \delta v \end{pmatrix} = -F(u,v), \tag{3.3}$$

at each stage to obtain better estimates $u + \delta u$ and $v + \delta v$. The quantities $F(u,v)$ and $\nabla_{u,v}F$ are straightforward to calculate as they arise directly from the explicit parameterisation of the template.

All that remains is the choice of $\mathbf{a}$ and $\mathbf{b}$. We obtain these vectors by taking the cross product of $\mathbf{w}$ with two arbitrary vectors — resulting in two vectors which are orthogonal to $\mathbf{v}$. More generally, the vectors $\mathbf{a}$ and $\mathbf{b}$ should be chosen to ensure that the system (3.3) is well-conditioned.

### 3.2 Updating the transformation estimate

The method we adopt to obtain an update of the transformation parameters is the Gauss-Newton method. This involves solving, at each iteration, the problem

$$J\delta\mathbf{t} = -\mathbf{d}, \tag{3.4}$$

in the $\ell_1$ sense, where $J$ is the Jacobian matrix of partial derivatives with entries $J_{ij} = \nabla_{t_j} d_i$. The estimate of the optimal transformation parameters is then updated according to

$$\mathbf{t} = \mathbf{t} + \delta\mathbf{t}.$$

Thus, since the distances $\mathbf{d}$ are obtained from the constrained closest point subproblem, we are left with the task of calculating the Jacobian matrix. For each datum, from equation (3.1), we have that

$$\widehat{\mathbf{x}}(\mathbf{t}) - \mathbf{f}(u(\mathbf{t}), v(\mathbf{t})) = \mathbf{w}d(u(\mathbf{t}), v(\mathbf{t})),$$

where we have explicitly included the dependency of the distance $d$ on the location parameters $U$. Differentiating and rearranging, we obtain

$$\nabla_{\mathbf{t}}\widehat{\mathbf{x}} = \mathbf{w}\nabla_{\mathbf{t}}d + \nabla_U \mathbf{f} \nabla_{\mathbf{t}} U.$$

This is equivalent to the form

$$\nabla_{\mathbf{t}}\widehat{\mathbf{x}} = [\mathbf{w} : \nabla_U \mathbf{f}] \begin{pmatrix} \nabla_{\mathbf{t}} d \\ \nabla_{\mathbf{t}} U \end{pmatrix}.$$

Therefore,

$$J \equiv \nabla_{\mathbf{t}} d = \mathbf{e}_1^{\mathrm{T}}[\mathbf{w} : \nabla_{\mathbf{u}}\mathbf{f}]^{-1}\nabla_{\mathbf{t}}\widehat{\mathbf{x}},$$

where $\mathbf{e}_1$ is the first component vector. Having obtained the Jacobian matrix $J$ and the distance vector $\mathbf{d}$, we are now in a position to solve the system (3.4) in order to update our estimate of the optimal transformation parameters $\mathbf{t}$.

We note that using the traditional orthogonal distances can lead to problems since calculation of the Jacobian matrix involves division of each row by the corresponding orthogonal distance — leading to exacerbation of rounding errors and possible division by zero especially in the $\ell_1$ case.

## 4 Numerical results

In this section, we present two example to illustrate the techniques presented in this paper. In the first, we have a small number of data which we wish to match to a given plane. In the second, we have a larger number of data and we wish to match them to a cylinder. In both cases, although analytical expressions are available to obtain the constrained closest points on the templates, we nonetheless utilise the method presented above in order to test its effectiveness.

### 4.1 Simple problem

Here we describe the problem of matching a representative set of 8 data onto the plane defined as

$$\mathbf{f}(u,v) = u\begin{pmatrix}1\\0\\0\end{pmatrix} + v\begin{pmatrix}0\\1\\0\end{pmatrix}.$$

Since this problem is rank deficient if we use all six possible transformation parameters, we restrict ourselves to using a translation in the $z$-direction and rotations about the $x$ and $y$ axes.

Having three degrees of freedom, we might expect to obtain an optimal $\ell_1$ solution which interpolates 3 of the data. However, as we shall see, this is unattainable in general and we can, in fact, only expect interpolation at two points. As Watson states [6], in such a situation, the rate of convergence can be unacceptably slow. This is found to be the case. It can be seen that not only is the convergence slow, but an optimal solution

| Iteration | norm(residuals) | norm(update) |
|---:|:---:|:---:|
| 1 | 0.6662 | 4.9901e-02 |
| 5 | 0.3008 | 3.5716e-04 |
| 10 | 0.3007 | 8.8545e-06 |
| 50 | 0.3006 | 9.1533e-06 |
| 100 | 0.3008 | 3.8514e-04 |

TAB. 1 Progress of the Gauss-Newton method for planar data.

is never obtained, with the objective function $\|\mathbf{d}\|_1$ increasing occasionally.

To ensure convergence, a simple line-search algorithm was adopted which searches along the direction obtained from the Gauss-Newton step for the maximum reduction in the objective function. This modification affects convergence in 3 iterations.

### 4.2 A more challenging problem

As a more challenging problem, we consider the matching of a set of 128 data which supposedly represent a cylinder but which contain 8 wild points. The cylinder is parametrised by $u$ and $v$ as

$$\mathbf{f}(u,v) = \begin{pmatrix} \cos u \\ \sin u \\ v \end{pmatrix},$$

resulting in a cylinder with unit radius oriented along the $z$-axis. Again, the problem of matching the data onto this model is rank deficient. The rank deficiencies occur due to rotations about the $z$-axis and translations along the $z$-axis. As such, we omit these possible transformations.

Although we might initially expect to interpolate 4 data points at an optimal $\ell_1$ solution, we find that in fact only two are guaranteed, although if a third point lies within two radii of one of these two points, then three points can be guaranteed. Typically, this will occur when the data is representative. For the data set we are considering, we expect three interpolation points due to the data representing the cylinder and in fact at the optimal solution, three interpolation points are obtained. In fact, the "missing" interpolation has the effect of slowing convergence of the Gauss-Newton method considerably so that in 100 iterations, the algorithm had not been deemed to converge. However, by the introduction of a simple line-search method, the algorithm converged in five iterations as displayed in Table 2.

| Iteration | norm(residuals) | norm(update) |
|---|---|---|
| 1 | 0.9654 | 5.6796e-03 |
| 2 | 0.9559 | 6.2932e-04 |
| 3 | 0.9557 | 1.0141e-04 |
| 4 | 0.9557 | 2.5812e-07 |
| 5 | 0.9557 | 4.4006e-14 |

TAB. 2 Progress of the Gauss-Newton method for cylindrical data using a line-search.

## 5 Conclusions

This paper has shown how perceived problems in $\ell_1$ template matching can be avoided by use of the so-called "method of directional constraints". In this method, the closest point on the template along a given direction vector is calculated in order to obtain the residuals between data and template. By then altering this direction vector to be the normal to the surface at that projected point, the algorithm progresses to the expected $\ell_1$ solution. Problems regarding undefined quotients are avoided by no longer updating

the direction vectors corresponding to a datum when the residual associated with that point is below a certain tolerance.

This work forms part of a larger project to consider novel approaches to ill-conditioned problems in metrology. It is hoped that the work presented in this paper will aid in the resolution of rank-deficient systems and ill-conditioned systems by altering the usual orthogonal distances to be these directional constraints, which should remove some of the rank deficiency.

As an example, consider the template matching problem where the template to be matched is an infinite cylinder with axis along the $z$-axis. Using typical template matching algorithms, this problem is rank deficient by two at the solution due to the possible translation in the $z$-axis and the possible rotation about the $z$-axis. By introducing these directional constraints, the rotational rank deficiency is almost completely resolved (there are now two possible rotations to obtain the optimal matching rather than the infinite number previously).

The use of the $\ell_1$ norm is also being used to attempt and resolve any rank deficiencies and ill-conditioning present in the problem. This is achieved by ensuring that any local deviations from the template (caused by, for example, wear) are "ignored" so that regions of local deviations might be compared. This will then result in a resolution of the uncertainty in the transformation parameters.

## Bibliography


1. B. P. Butler, A. B. Forbes, and P. M. Harris. Algorithms for geometric tolerance assessment. Technical Report DITC 228/94, National Physical Laboratory, Teddington, UK, 1994.
2. V. Jovanovski. Three-dimensional Imaging and Analysis of the Morpohology of Oral Structures from Co-ordinate Data. Ph.D. Thesis, Department of Conservative Dentistry, St Bartholomew's and the Royal London, School of Computing and Dentistry, Queen Mary and Westfield College, London, UK, 1999.
3. D. A. Turner. The approximation of Cartesian coordinate data by parametric orthogonal distance regression. Ph.D. Thesis, School of Computing and Mathematics, University of Huddersfield, UK, 1999.
4. D. A. Turner. Least squares profile matching using directional constraints. Preprint, 2001.
5. G. A. Watson. *Approximation Theory and Numerical Methods.* Wiley, New York, US, 1980.
6. G. A. Watson. On curve and surface fitting by minimizing the $\ell_1$ norm of orthogonal distances. Preprint.
7. D. S. Zwick. A planar minimax algorithm for analysis of coordinate measurements Advances in Computational Mathematics, **2**:4, 1994, 375–391.